Job Access and Property Price Gradients Over Time

Nathan Wiseman

June 1, 2017

Abstract

Demand for housing should be closely linked to stable income earning opportunities. For this reason, spatial dynamics of house prices are likely to change with changes in job accessibility. This paper aims to contribute to the literature on this topic by providing estimates of the impact of changes in job accessibility over time and at a finer geographic scale. In contrast to previous studies that have found no relationship between changes in job access and changes in home prices, it is found that there are significant effects that vary by sign and magnitude depending on job type. The results have implications for estimating the economic impacts of firm location decisions and the impact of the business cycle on the spatial distribution of housing prices.

1 Introduction

Price gradients are a measure of how prices change over physical space. The first to study land prices and their relation to space was von Thunen [1826], who developed the theory of locational rent. This theory states that land values are highest inside a city, which is surrounded by a series of rings of lower value uses such as fuel, crops, and ranches. Alonso [1964] built on this idea with the "bid-rent curve," where those with the most productive uses of the land would bid up prices nearest the Central Business District (CBD). By locating in a centralized area, businesses were better able to minimize transportation costs to the market. In addition, they had access to a larger labor market, a greater variety of capital inputs, and more support industries. These benefits are the basis for agglomeration effects (Marshall [1920]; Krugman [1991]; Glaeser et al. [1992]), where clusters of businesses develop in the same vicinity.

As a result of many cities sprawling out into the suburbs in the 1950s and 1960s, the central business district has become less important to a city's production. Econometric studies over time have begun to show evidence that the traditional theories of price gradients may hold less water than they once did, with studies showing that prices in many cities actually decline as homes become closer to the CBD. In many cases, there is evidence of competing "sub-centers" developing, which, due to their relatively lower land costs, have attracted firms to locate outside the CBD. In addition to lower land costs in the suburbs, the increasing congestion of major metropolitan areas has created incentives for firms and workers to locate in new areas. Over time, econometricians developed ways to account for multiple employment centers, employing techniques to locate sub-centers and measure the impact of accessibility to employment in multiple areas. In recent years, there has been evidence that employment has become even more dispersed in many large cities, and more sophisticated measures of job accessibility have been developed in order to better understand the spatial relationship between employer locations and where people live.

While there is a long line of literature on property prices and their relationship to job accessibility, less is known about the dynamic relationship between the two. The few studies that have attempted to shed light on the impact of changes in job accessibility over time have found inconclusive results. This paper makes several contributions to the literature by estimating the impact of changes in different types of employment on the appreciation rates of properties over time. The next section will review the literature on these topics. This is followed by a discussion of the data and models used for the analysis. Results of the study follow, which suggest that changes in employment can have substantially different impacts on property prices over time, depending on the type of employer. These results could provide valuable insights for practitioners that aim to develop strategies for regional economic development.

2 Literature Review

The first measures of job accessibility were based on the assumptions of a monocentric rent gradient centered at the Central Business District (CBD). Initial attempts tended to use linear distances, but Hansen [1959] used a negative exponential function on travel costs. More recently, measures that allow for polycentric cities with multiple employment centers have been developed. Giuliano and Small [1991] provide a methodology for identifying employment centers, and many studies have applied this methodology (Song [1994], Small Kenneth et al. [1994], McMillen and McDonald [1998]). Others have since developed other ways to identify sub-centers in a polycentric city (e.g. McMillen [2003a]). Nelson [1977] used the travel time to reach a specified percentage of the region's employment, while Burnell [1985] looked at manufacturing jobs weighted by the inverse distance. Song [1996] compares 9 different accessibility measures in the context of population density. The rest of this section will review some important findings regarding price gradients, with a focus on more recent studies. Look to Yiu and Tam [2004a] for a more thorough review of studies prior to 2004.

Lee [2007] looks at changes in job dispersion over time in different MSAs using geographically weighted regression in addition to traditional methods of identifying employment sub-centers. The data comes from the 1980, 1990, and 2000 census journey-to-work surveys, and gives employment densities at the census tract level. The main findings of the study are that workplaces are becoming considerably more dispersed than would be expected from the polycentric assumption. However, it is also found that there are significant differences between cities, with the CBDs in New York and Boston tending to remain the dominant employment center, and employment sub-centers becoming more pronounced in polycentric cities such as San Francisco and Los Angeles.

Franklin and Waddell [2003] use data for King's County, WA from 1995 and 1998 and divide employment into commercial (retail, office, and government), educational, and industrial (manufacturing, warehousing, communications, transport, and utilities). In addition to looking at different types of employment, they use congested travel times as their measure of distance. They compute elasticity with respect to each activity type, finding that commercial and university employment have positive effects on home prices, while educational and industrial employment have negative effects.

Matthews [2006] distinguishes between pre- and post-war neighborhoods, noting that the impact of accessibility differs between pre-war, grid-style developments, and post-war, suburban neighborhoods with cul-de-sacs. Their main contribution is recognition that while accessibility is valuable, there are negative externalities associated with close proximity due to increased traffic, noise pollution, and obstruction of views. They find that in the grid-style neighborhoods where people are more likely to walk or use transit, positive externalities dominate once establishments are more than 250 feet away; whereas, for post-war neighborhoods that are more car oriented, the negative externalities always dominate. Their analysis is based on 38 census tracts in two areas of Seattle.

Several other authors have noted the non-monotonic relationship between accessibility to employment and housing prices. Osland and Pryce [2012] give a good review of the inconsistencies of the land rent gradient over time (for early evidence of this phenomenon, see Clapp et al. [2001]), and suggest that the source could be a combination of polycentric employment centers and non-monotonic rent price gradients due to negative externalities closer to employment centers. They construct a gravity based model that allows for non-monotonicity and multiple employment centers. They estimate their model using spatial econometric techniques on a cross-section of homes sold in Glasgow, Scotland in 2007. Using employment data for 6501 employment zones , they find evidence of non-monotonicity. Some limitations of their study, due to data limitations, are that they look at aggregate employment rather than employment by sector and use Euclidean distances in their accessibility measure rather than travel times. Gibb et al. [2014], in a follow-up paper using the same data and methodology, estimate the monetary value of employment access.

Ottensmann et al. [2008] repeats Song's analysis of accessibility measures with a focus on housing prices. They use a travel demand model to calculate travel times between Traffic Analysis Zones (TAZs). Most recent studies have abandoned the approaches based on the monocentric or polycentric assumptions, instead opting for measures that do not require specification of the peaks of the gradient and that impose less stringent limits on the functional relationship. For example, Cerda and El-Geneidy [2009] use 10 minute and 30 minute isochrones, while Giuliano et al. [2010] looks at number of workplaces by sector, and Iacono and Levinson [2011] use 30 minute isochrones.

Hwang and Thill [2010] use census tract data for Seattle & Buffalo. They perform a factor analysis and show that the impact of job accessibility varies spatially using geographically weighted regression.

As can be seen from the review of studies above, there is ample evidence that there is a relationship between the spatial distribution of home prices and accessibility to employment. A major limitation of these studies is that they are all based on cross-sectional data. This makes it difficult to determine whether there is a causal relationship, and, if so, whether differences in accessibility are responsible for the observed patterns.

Only a few authors have considered the dynamic relationship between home prices and job accessibility. Accessibility is likely to change in response to firm location decisions, increases in transportation capacity, and changes in congestion patterns associated with population growth. McMillen [2003b] applied a Fourier repeat sales approach to estimate locally weighted repeat sales models for each census tract in Chicago between 1983 and 1998. The Fourier approach, by imposing smoothness on the price gradient, allowed for estimation of price indices for census tracts which would have otherwise not had enough repeat sales. Subsequent regressions of the appreciation rates on demographic factors and distance to the CBD indicate that homes nearest the CBD appreciated the most, though the kernel density estimates show the relationship to be non-monotonic. Yiu and Tam [2004b] study the impact of accessibility along a railway in Hong Kong between 1994 and 2001 using repeat sales and hedonic index formulations. With the repeat sales index, they maintain the monocentric assumption, allowing distance to the CBD to affect appreciation rates. With the hedonic index, they interact time variables with location dummies in order to estimate separate price indices at 8 different stations.

While McMillen [2003c] and Yiu and Tam [2004b] are able to identify differential appreciation rates based on geographic location, they do not explicitly link these appreciation rates to changes in the number of jobs. Iacono and Levinson [2017] use two years of data (2000 and 2005) to look at the impact of changes in job accessibility on price changes. They use a novel approach, defining a "representative house" in each traffic analysis zone based on the median price and characteristics of homes sold in that area in that year. Their main finding is that while levels of job accessibility are correlated with home prices, changes in job accessibility are not significant predictors of changes in home prices, calling into question the causal nature of the relationship. Nonetheless, they note several limitations to their approach that could affect the validity of their results. They suggest that future studies consider using a longer time frame along with a repeat sales methodology instead of the "representative house" in order to better control for structural differences and local amenities that are time-invariant.

A summary of some important findings of past studies are as follows:

1. The impact of job accessibility could differ depending on the type of employment (Franklin and Waddell [2003]).
2. The measurement of distance to employment is important, with measurements based on time rather than distance to be preferred (Ottensmann et al. [2008]).
3. Traditional accessibility measures may lead to inconsistent results if there are non-monotonic price gradients due to negative externalities (Osland and Pryce [2012]).
4. Spatial patterns of employment are changing over time within urban areas, with trends towards multiple employment centers and more dispersion (Lee [2007]).
5. There is limited evidence on the relationship between changes in accessibility and changes in home prices (Iacono and Levinson [2017]).

This paper builds on these findings in an attempt to better understand the relationship of accessibility to employment and the housing cycle. First, accessibility measures are defined at a much higher resolution, with employment counts at the census block level by industry sector. Second, accessibility is measured using transportation times at small intervals in order to assess the non-monotonic nature of the relationship. Finally, the analysis is the first to consider the dynamic relationship between job accessibility and home prices in a repeat sales specification without a monocentric or polycentric assumption. This allows for the impact of job accessibility to vary over time in complex ways, while also having the benefit of being able to control for time invariant factors associated with each home.

3 Data

Data on home prices was obtained from the Washoe County Assessor's web page. This dataset contains the most recent tax assessment for all properties, as well as sales data from the county recorder. While the data includes characteristics, such as the number of bedrooms and bathrooms, lot size, gross living area, garage and basement size, fireplace, porch, etc., the repeat sales methodology is parsimonious, in that it only requires information on sale prices and dates. The analysis in this paper considers only fair-market sales of single family residences. There were a total of 95,031 single family residences in the assessor data. These properties account for 78,560 sales within the period of interest (2002-2014). Of these sales, 15,315 sales constitute repeat sales.

The street address of the property was used to obtain latitude and longitude coordinates through the census geolocator service API, which also returned the census block FIPS code for the property. The service was able to successfully geolocate 85.4% of the properties. Upon review of the data, it was found that successive assessor's parcel numbers (APNs) tend to be in the same census block. Thus, a nearest neighbor algorithm, based on APN, was used to assign the remaining houses to census blocks.

Employment data comes from the Census Longitudinal Employer-Household Dynamics (LEHD) survey. Specifically, the data comes from the LEHD Origin-Destination Employment Statistics (LODES). The LODES data provides worker counts by 3 digit NAICS code and demographic characteristics at the census block level annually from 2002 to 2014.

Travel times were calculated using Open Trip Planner (OTP), an open source route planning program. The OTP program was used to calculate the area reachable within 5 minute intervals between 5 and 40 minutes from the centroid of each census block. The spatial polygons, known as isochrones, where used to determine which census blocks were accessible within the given time frame. If a census block's centroid was located within the calculated isochrone, it was determined to be reachable within that amount of time.

Following the methodology used in the EPA Smart Location Database, the NAICS sectors were aggregated into 5 categories: retail, industrial, office, entertainment, and service. The amount of jobs in each category that were reachable within the time intervals was calculated for each year by matching the reachable blocks with the LODES data, and the results were merged to each property based on the property's census block and year of sale. The final result was eight variables for each property for each employment category.

Figure 1: **Average Accessibility by Travel Time**

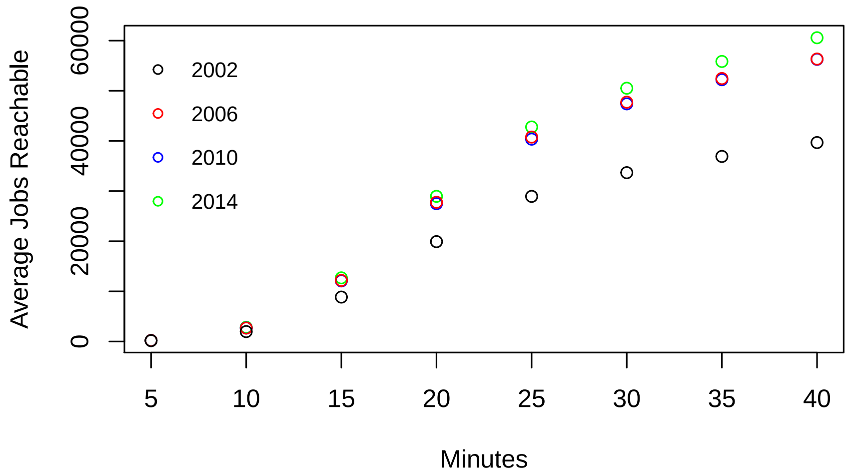


Figure 1 shows the average number of jobs within a given travel time for 4 different years within the study period. Between 2002 and 2006, there were large changes in the total number of jobs accessible from the average home. Most of these changes appear to have occurred between 15 and 30 minutes from the average home. Between 2006 and 2010, the average number of jobs reachable remained approximately constant, and there was a small increase between 2010 and 2014.

Figure 2: **Jobs Accessible Within 15 Minutes by Area. LEFT: Accessibility in 2002. RIGHT: Accessibility in 2014.**

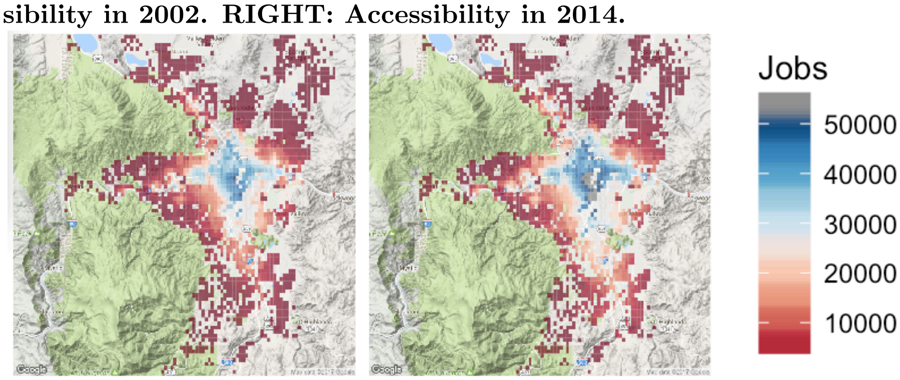


Figure 2 shows two maps of the study area. Overlaid on each of the maps is the number of jobs paying more than $3333 reachable within 15 minutes travel time. The map on the left shows the number of jobs reachable in 2002, while the map on the right shows the number of jobs reachable in 2014. As can be seen, the number of jobs reachable is highest for homes located nearest the Central Business District (CBD) in both years; however, it is also apparent that there were substantial increases in accessibility for most homes over this period. This measure, based only on jobs paying more than $3333 or more a month, was chosen because it is unlikely that workers making less than this amount would be able to qualify for a mortgage. Thus, focusing on changes in these types of jobs could better reveal relationships that other studies have not been able to find. Due to large amounts of collinearity between the number of jobs reachable within different travel times, the accessibility measure used in this study is the difference in the number of jobs accessible between 5 minute intervals. For example, separate variables were constructed to measure the amount of jobs within 5 minutes, between 5 and 10 minutes, between 10 and 15 minutes, etc., up to 40 minutes. Thus, the variables are similar to concentric rings emanating from the block group each household is located in, only they are of varying shapes due to differing transportation infrastructure and natural barriers.

4 Models

This section describes two different model specifications that are common in the literature on explaining home prices and building price indices. A thorough discussion of the relationships of the two approaches and the assumptions required for them to be unbiased and consistent estimators is provided. In addition, an important hypothesis is formulated, which may be able to help understand whether existing implementations of repeat sales models are appropriate.

4.1 Hedonic Regression Models for Home Prices

There is a close relationship between the repeat sales approach and the hedonic regression model. Homes are comprised of a variety of characteristics, which can in general be thought of as several factors that contribute to the value of the home, such as its size, location, quality, and condition. These factors are obviously considered by a home buyer when they are forming their willingness to pay, but the contributions of each are not directly measurable, as they are not transacted in their own market places due to their inseparable nature. For this reason, a home purchase can be thought of as a bundle of characteristics, each of which contribute a certain amount of value to the sale price realized in the market. The goal of a hedonic regression is to estimate the unobserved dollar value of each of the characteristics. In mathematical terms, the hedonic model can be represented as follows:

$$Price = f(SIZE, LOCATION, QUALITY, CONDITION)$$

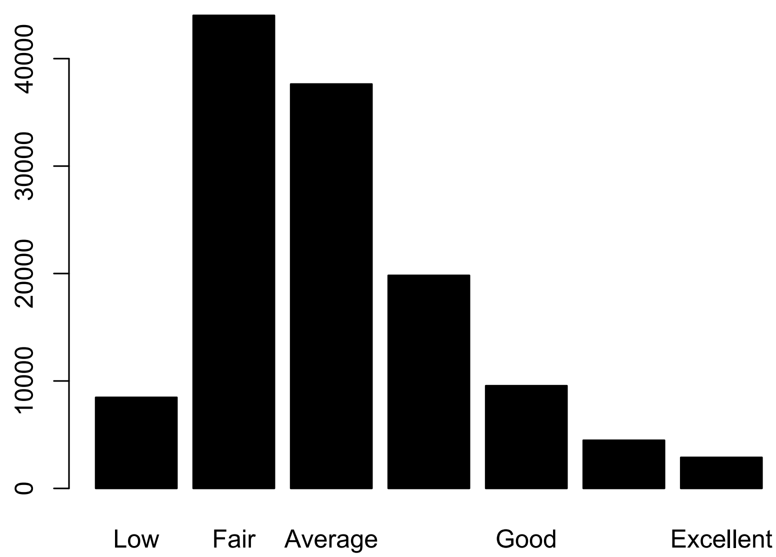
where the factors $SIZE$, $LOCATION$, $QUALITY$, and $CONDITION$ are themselves comprised of several dimensions, which may or may not be directly measurable. In the discussion that follows, the terms feature, variable, and measure are used interchangeably to mean an attribute or characteristic that data is available on, which one might try to use directly or through a transformation such as logarithms, square roots, or polynomial expansions, in order to decompose the sale price into the individual contributions of these characteristics. Whether a transformation is necessary is often determined through a combination of theory and empirical analysis of relationships between the observed variables in a given sample. For example, it might make theoretical sense to test the explanatory power of nonlinear functions of age. Age is unique in that it is associated with all of the other factors due to the relationship of these factors over time. For example, it is likely that older neighborhoods will have smaller homes located closer to the central business district (CBD). In addition, construction techniques that affect quality could vary over time, with historic homes that are still in use being more likely to be built of brick or stone. Finally, condition has a complex relationship with age and quality, since over time homes may depreciate at different rates depending on the quality of the construction materials and the level of maintenance performed on the home. Below is a discussion of the four main factors in the equation above, and the features commonly used as measurements or proxies for them.

Measures of size could include the size of the land, the square footage of the home, the number of bedrooms and bathrooms, the number of stories, and whether it has utility rooms such as a garage or basement that are not included in the square footage of the home.

Variables commonly used to control for location include data on factors that vary over space, such as school quality and accessibility to jobs and other amenities such as parks and recreation. Another location specific feature that is more difficult to control for is view. Jurisdictional boundaries that determine the level of taxes, public amenities, and access to utility providers such as water, electricity, and sewer, are other possible variables that could change over space. One common approach to control for factors that vary over space is to proxy for the features described above by including indicator variables for different neighborhoods that enjoy similar levels of these variables or fall within the same jurisdictions. Often these measures of location are based on census geographies such as census tracts, block groups, or blocks.

Quality of the home is often measured by features such as the materials used for the siding (e.g., brick, stucco, wood, masonry, etc.) and roof (e.g. wood or shale), whether the home has non-standard features such as a pool or fireplace, and other features that may be standard in certain regions, such as the presence and/or type of air conditioning and heating. In Washoe County, the assessors provide a quality class for the homes, with the quality class ranging from poor to excellent and several rare very high quality classes. This seems to be a unique feature of the data used in this study area, as a review of the literature for similar variables does not turn up mention of such a measure. The distribution of the reported quality of the homes in the sample are shown in Figure 3.[1]

Figure 3: **Distribution of Quality Class**



As can be seen, fair and average homes are the most common in the housing stock, but there is a long right skew, with a small number of homes being ranked as good, very good, or excellent.

Figure 4 shows the percent of each quality class by the year the home was built. Homes of fair quality routinely made up more than half of the housing stock prior to 1990. After 1990 homes of average quality were the most common additions to the housing stock. Around 2003, when the recent housing bubble was expanding, homes of above average quality began to be built more rapidly. While the models developed in the next section are applied to data from 2002 to 2014 (since 2014 is the last year available for the LODES data), the most recent data on construction quality is particularly telling: so far in 2017, 77% of all homes were of average or above average quality, with only 13% having below average quality.

[1]The Washoe Count Assessor includes 14 possible quality classes; however, many of them are somewhat uncommon. For example, there are 5 "High Class" levels of quality that constitute only a few percent of the overall housing stock. Furthermore, there are several intermediate classes reported, such as "fair-average" which were combined with the class below.

Figure 4: **Percent in Each Quality Class by Year Built**

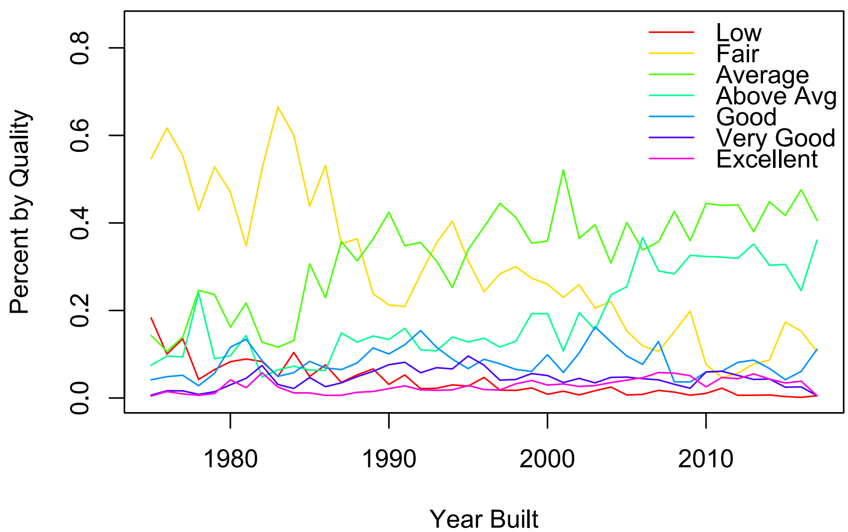
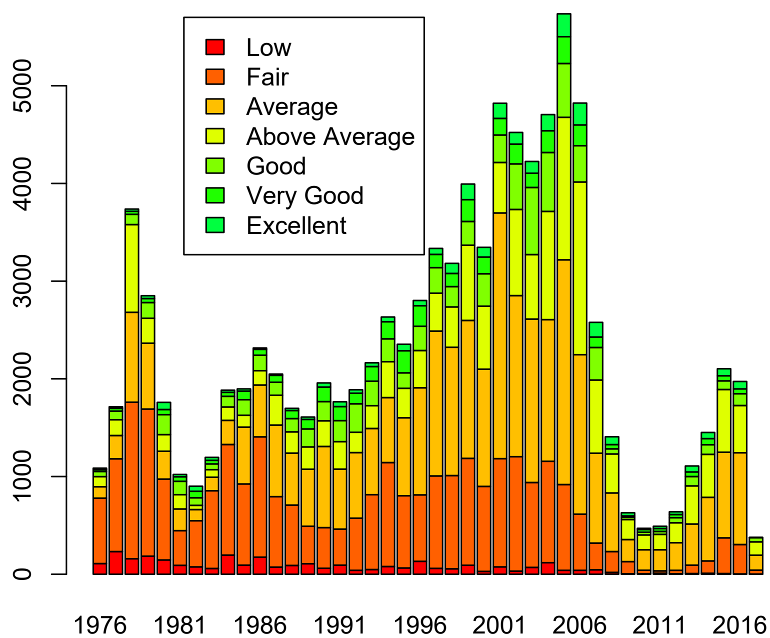


Figure 5 illustrates the number built per year by quality. The most striking thing about this graph is the massive decline in the number of houses being built in Washoe County since the housing bubble. While the pre-2008 average was close to 3000 new homes per year, construction fell to below 1000 units a year until recently.

Another factor that is closely related to quality is condition. Condition is perhaps the most difficult to measure directly, but a commonly used proxy is the age of the dwelling or the year it was built.

Figure 5: **Number in Each Quality Class by Year Built**



The basic equation of a hedonic model, which assumes that the factors are additive in the characteristics that contribute to their value, is of the form:

$$Price = \alpha + \beta X + \epsilon$$

where X is a vector of the characteristics that are expected to influence the price of the dwelling, ϵ is a stochastic error term, α is an intercept parameter that is to be estimated, and β is a vector of slope parameters to be estimated, which measures the contribution of each of the characteristics. The value of the hedonic model is its ability to estimate the dollar weights on each of the characteristics included in the model. This is typically done by applying linear regression of the price of the home on the homes characteristics.

Of course, the ability to obtain unbiased and consistent estimates of the effects of each characteristic on the final price depends upon the traditional assumptions of the OLS estimator. First, it is required that the functional form is representative of how people actually value the characteristics, including any interactions between characteristics or non-linearity. Violation of the first assumption will tend to result in what is known as misspecification bias. Second, it is required that the stochastic error term ϵ is not correlated with any of the variables included in the regression. If variables that influence price and are correlated with the available characteristics are not measurable, this will result in a violation of the second assumption and induce bias in the estimated parameters due to endogeneity. Third, it is required that the homes come from a random sample of the housing stock, which ensures that the value ranges and characteristics are representative of all homes. If homes with certain characteristics are more likely to be sold, this would violate the third assumption and create what is known as selection bias.

All of these types of bias will tend to result in inconsistent coefficient estimates, meaning that even with large samples it will not be possible to identify the true value placed on each of the characteristics. As discussed above, there are many relevant characteristics that could contribute to the final value of a given home. With so many characteristics, there are many possible combinations of variables, their interactions, and transformations that would need to be considered in order to obtain the correct functional form. Furthermore, since it is likely that there are factors related to location, quality, and condition that are not measured or are measured with error, it is likely that most hedonic regressions suffer from endogeneity. Finally, it is likely that preferences, for example for a given location or size of home, would change over time. This would imply that homes with certain characteristics would be more likely to transact in a given period, leading to selection bias.

4.2 Repeat Sales Index

Clearly, there are many methodological concerns that make interpretation of the coefficients estimated by the hedonic method unlikely to be truly representative of the values placed on each characteristic. Another approach that is commonly employed, known as the repeat sales method, is able to address some of these concerns. Following McMillen [2012], consider the more general specification of the hedonic model:

$$Price = \alpha + \beta X + \gamma W + U$$

where X, α, and β are interpreted as in the traditional model described above. This equation includes an extra set of variables, W, which are unmeasurable, but affect the price of homes. Assume that this set of variables is comprehensive such that the new stochastic error term, U, representing the unexplained variation in prices after controlling for the sets of variables X and W, is not correlated with any of the regressors. Furthermore, assume that these sets of variables contain any appropriate interactions and transformations, such that the model has the correct specification, and that the homes that have sold are a random sample of the housing stock. In other words, assume that assumptions one and three from the discussion above are satisfied. Under these circumstances, an OLS regression would still be biased if the variables contained in W were correlated with any of the variables contained in X. McMillen [2012] shows how the traditional repeat sales estimator can be derived from the hedonic equation for homes that sell multiple times. Essentially, the hedonic regression above could be estimated for each period, and one could investigate changes in prices rather than their levels. For illustration, consider a simple model containing only two time periods for which the sample of homes that sell in both time periods is representative of the overall stock of housing. Subtracting the hedonic equation of the first period from the hedonic equation of the second period results in the following equation:

$$Price_{t'} - Price_t = (\alpha_{t'} - \alpha_t) + (\beta_{t'} X_{t'} - \beta_t X_t) + (\gamma_{t'} W_{t'} - \gamma_t W_t) + (U_{t'} - U_t)$$

Under certain assumptions, namely that the value of the characteristics in X and W are not changing over time, and that their respective values (the vector of coefficients β and γ) are constant, the equation reduces to:

$$Price_{t'} - Price_t = (\alpha_{t'} - \alpha_t) + (U_{t'} - U_t)$$

Since it is likely that the coefficients of the characteristics would increase when the price increases, the assumptions required to drop the terms involving X and W are unlikely to be satisfied. Typically, repeat sales regressions are estimated using the log price as the dependent variable.[2] With this specification the assumption is slightly less strict. Since converting to logs makes the coefficients interpretable as percentage changes in price due to a one unit change in the characteristic, this specification assumes that these percentage changes are constant rather than the dollar contribution, which allows an increase in a characteristic to have a different impact depending on the initial price of the home, or, equivalently, depending on the levels of the other explanatory variables. Essentially, by applying this assumption over a longer time frame, it is possible to generalize this model to any number of periods.

Traditionally, repeat sales indices only include variables that provide information about the price of homes that have sold multiple times. The model structure, as introduced by Bailey et al. [1963], is shown below:

$$\ln(\frac{P_{it'}}{P_{it}}) = \alpha_{t_1} D_{it_1} + \alpha_{t_1} D_{it_1} + ... + \alpha_{t_n} D_{it_n} + V_{it'}$$

where $\ln(\frac{P_{it'}}{P_{it}})$ is the natural log of the ratio of the observed sale prices for a home that sold first in time t and sold again in time t', and $t, t' \in (t_1, t_n)$. The variable D_{it} contains information on whether the home sold in period t. If a home did not sell in the time period, D will be coded as 0; otherwise, the value of D depends on whether the sale that occurred in time t was the first or second

[2]Using the log-linear functional form implies that the true model for prices is of the form $Price = e^{\alpha} e^{\beta X} e^{\gamma W} e^{U} = e^{\alpha + \beta X + \gamma W + U}$.

sale of the pair, where it is coded as a -1 if the first sale occurred and a 1 if the second sale occurred. It is straightforward to show that this is equivalent to the equation obtained above based on the assumption of equal coefficients and values of the explanatory variables for any two periods, where $V_{it'} = (U_{t'} - U_t)$.

The derivation of the repeat sales model from the hedonic model is useful in understanding the implicit assumptions of the repeat sales model, and in illustrating how using the repeat sales methodology removes the complexity associated with model specification and endogeneity from omitted variables contained in W. Nonetheless, it is still subject to selection bias if the homes sold in each period are not representative of the housing stock, and the new assumptions that allow for the variables contained in X and W to cancel out will be violated if these variables and/or their coefficients are changing over time. Several studies (Gatzlaff and Haurin [1997], Gatzlaff and Haurin [1998], Jud and Seaks [1994]) have applied methods for selection bias corrections in both the hedonic and repeat sales models based on the Heckman [1977] two-step correction. In order to provide tractability, the remaining of this paper will assume that the assumption that sales in each period are a random sample of the population is valid.[3] Other studies have developed hybrid approaches in an attempt to control for the characteristics that vary over time. For example, McMillen and Thorsnes [2006] controls for changes in quality using building permit information.

Despite acceptance of job accessibility as an important characteristic of homes in the hedonic model, no known studies have looked at changes to job accessibility in the context of a repeat sales model. Since job accessibility is likely to change with economic conditions, it is unlikely that the traditional repeat sales model with only time variables would be specified correctly. Nonetheless, it is possible to control for changes in the values of the variables in X as long as the coefficients remain constant. Specifically, the regression equation of interest is:

$$\ln\left(\frac{P_{it'}}{P_{it}}\right) = \sum_t \alpha_t D_{it} + \sum_a \beta_{Access\Delta_a}(Access_{iat'} - Access_{iat}) + V_{it'}$$

where a corresponds to the type of job access (i.e., industrial, retail, etc.). Essentially, if the coefficient on job accessibility is significant in this specification, it indicates that the traditional repeat sales index is not specified correctly due to the invalid assumption that all relevant variables are not changing over time. This is the relevant hypothesis test in this study:

$$H_0 : \beta_{Access\Delta_a} = 0 \quad \forall \quad a$$

Rejecting the null hypothesis would provide evidence that the traditional repeat sales index would be biased, providing support for the use of the hybrid model that is able to control for characteristics that change over time.

[3] For a more thorough review on the evidence of selection bias, a critique of its application, and a potential new method to correct for selection bias, the reader is referred to Wiseman [2017].

5 Results

Table 1 compares the results from the traditional repeat sales method, using only time period variables to account for changes in prices, to the hybrid model that separates out the impact of changes in accessibility. In both regressions, the majority of the coefficients are highly significant; however, there are large differences, particularly in the bubble years from 2003 to 2009. Furthermore, as can be seen in Table 2, the majority of the accessibility measures are also significant, but with different signs. Both models suffer from heteroskedasticity, as is known to be an issue due to different holding periods (Case & Shiller, 1989). The heteroskedasticity was corrected for using the Case & Shiller (1989) method.

The results suggest that different types of employment have different impacts on home price appreciation rates, and that the impact is not always decreasing with travel time. Of note, changes in retail employment would be expected to increase the appreciation rates of homes within a 15 minute radius, with the largest impacts on homes within 5 minutes. The significant coefficient on $Retail\Delta$ for 5 minutes suggests that a retail establishment that would hire 10 employees would increase the value of homes within 5 minutes by approximately 1%. The coefficient on $Entertainment\Delta$ for 5 minutes suggests a similar effect. The coefficient on $Retail\Delta$ for 10 minutes suggests that homes between 5 and 10 minutes from the establishment would see a little less than half the benefit, while the coefficient on $Entertainment\Delta$ for 10 minutes suggests there is no significant benefit of entertainment access once you are 5 to 10 minutes away from it.

Table 1: **Repeat Sales Regression Time Index Results. Model 1 is the traditional Case and Shiller [1989] repeat sales index. Model 2 includes changes in accessibility (See Table 2 for the coefficients on the accessibility measures).**

	Dependent variable:	
	$\log(\frac{Price_{t'}}{Price_t})$	
	(1)	(2)
y2002	-0.286^{***}	-0.621^{***}
	(0.012)	(0.033)
y2003	-0.127^{***}	-0.493^{***}
	(0.011)	(0.033)
y2004	0.121^{***}	-0.175^{***}
	(0.011)	(0.026)
y2005	0.386^{***}	0.074^{***}
	(0.011)	(0.025)
y2006	0.396^{***}	0.096^{***}
	(0.011)	(0.033)
y2007	0.310^{***}	0.045^{*}
	(0.012)	(0.027)
y2008	0.135^{***}	-0.034
	(0.013)	(0.027)
y2009	-0.175^{***}	-0.322^{***}
	(0.013)	(0.025)
y2010	-0.276^{***}	-0.354^{***}
	(0.012)	(0.025)
y2011	-0.385^{***}	-0.408^{***}
	(0.013)	(0.027)
y2012	-0.353^{***}	-0.429^{***}
	(0.012)	(0.021)
y2013	-0.154^{***}	-0.199^{***}
	(0.012)	(0.015)
Observations	13,030	13,030
R^2	0.607	0.623
Adjusted R^2	0.607	0.622
Residual Std. Error	0.352 (df = 13018)	0.345 (df = 12978)
F Statistic	$1{,}674.892^{***}$ (df = 12; 13018)	413.128^{***} (df = 52; 12978)
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2: **Repeat Sales Regression Accessibility Coefficients (reported as semi-elasticities). The accessibility measure used is based on the additional number of jobs paying more than $3333/mo that were reachable in successive 5 minute intervals. For example, the coefficient for $Service\Delta$ with travel time of 15 minutes is the percentage change in home prices due to an additional job in the service sector located between 10 and 15 minutes from a home.**

Travel Time	$Retail\Delta$	$Industrial\Delta$	$Office\Delta$	$Entertainment\Delta$	$Service\Delta$
5	0.10577	0.02498	0.04553	0.10978	-0.02349
	(0.02984)***	(0.01453)*	(0.01962)**	(0.04294)**	(0.00732)***
10	0.0407	0.00315	-0.01095	-0.009	-0.00693
	(0.01201)***	(0.00286)	(0.00371)***	(0.00958)	(0.00117)***
15	0.01866	0.0034	0.00459	-0.00587	-0.00492
	(0.00943)*	(0.00118)***	(0.0019)**	(0.00771)	(0.00077)***
20	0.00868	-0.00005	0.00379	-0.0068	-0.00364
	(0.00802)	(0.00107)	(0.00177)**	(0.00739)	(0.00069)***
25	0.00167	0.00117	0.00246	-0.01449	-0.00313
	(0.00867)	(0.00106)	(0.00163)	(0.00747)*	(0.00068)***
30	-0.00937	0.00479	0.0037	-0.02203	-0.00322
	(0.00862)	(0.00114)***	(0.00174)**	(0.00775)***	(0.00068)***
35	0.00454	0.00546	0.00635	-0.01247	-0.00328
	(0.0081)	(0.00131)***	(0.00139)***	(0.00762)	(0.00078)***
40	-0.01285	0.00056	0.00624	-0.02096	-0.00216
	(0.00661)*	(0.00135)	(0.00154)***	(0.00718)***	(0.00084)**

The results also suggest that office jobs have an overall positive impact (other than the small, but significant negative coefficient for changes in office employment within a 5 to 10 minute travel time). Surprisingly, the results show that service jobs have a negative impact for all distances considered, with the impact larger the closer the jobs. Previous literature has suggested that industrial jobs may be most likely to create negative externalities; however, there are no significant negative coefficients for industrial jobs. In fact, several of the coefficients on changes in industrial jobs are actually positive, with the biggest positive impact coming for jobs 25 to 35 miles away. It is likely that this is catching the impact of the Tahoe-Reno Industrial Center, which has become a major center for industry in nearby Story County. Recently, Tesla announced that their Gigafactory is likely to employ more than 10000 jobs. The results of this study suggest that if all of those jobs pay more than \$3333 /mo, this could have a large impact on home prices in Washoe County, as the coefficients on $Industrial\Delta$ suggest homes between 25 and 35 minutes from the plan could expect to see increases of around 50%.

One thing that is somewhat unusual about the jobs coming to the area from the Gigafactory is that the market has already had almost 3 years to adjust to the news. While the efficient markets hypothesis, which states that prices account for all publicly available information, is thought to hold true for more liquid assets such as stocks (Fama [1965]), the market for housing is thought to be less efficient due to large transaction costs in selling and building homes. Home prices in Washoe County have increased over 30% since the factory was announced, and with recent news that Google just purchased land next to the Tesla plant at the Tahoe-Reno Industrial Complex, it is likely that homes within the vicinity will continue to appreciate.

6 Conclusion and Future Work

Previous studies have not found significant impacts of changes to accessibility on changes in home prices (Iacono and Levinson [2017]). The results of this paper suggest that changes in job accessibility do in fact affect home price appreciation rates. There are several possible explanations for these findings. First, by utilizing a repeat sales index methodology, this study was able to account for changes in accessibility over longer periods of time and while controlling for unobserved household fixed effects. Furthermore, the patterns found for different employment types reiterate the importance of considering non-monotonicity and disaggregating employment. Finally, this study was done at a much finer geographical resolution, using census blocks rather than traffic analysis zones and census tracts, which have been the basis of previous studies. This has allowed for more accurate measurements of accessibility and for a better understanding of how the impact of access varies by distance. Finally, the measure of accessibility utilized in this study was based on jobs that pay more than \$3333 per month, which could also be a significant factor in the results that were obtained. Since these jobs are more likely to provide enough income for the employees to qualify for a mortgage, they are more likely to have a significant impact on housing demand.

One limitation of the study is that the accessibility measures were all built using the current infrastructure. It is possible that this is causing some measurement error, particularly for homes located towards the south end of Washoe Valley, where the two-lane road that used to connect Reno to Carson City was replaced with Interstate 580 in 2012. This major infrastructure change reduced the driving time from approximately 35 to 25 minutes between the edges of the two cities, which is likely to have had a significant impact on the accessibility for some residents. Many authors have used major infrastructure changes to estimate changes in accessibility (see, e.g., Smersh and Smith [2000]).

Another potential issue is that, while the study period was longer than previous studies on the topic, it is still a short period in which to consider a repeat sales model. Overall, the majority of repeat sales had to be eliminated from the sample since the accessibility measure was only available between 2002 and 2014. Home sales in general are perhaps more likely to suffer from sample selection bias, and previous studies have suggested that the repeat sales methodology could exacerbate this problem. By considering such a short period, the problem could be further compounded if the homes found to sell more than once in the sample are not representative of the overall housing stock.

Future work in this area could focus on addressing the methodological issues discussed above, and could also consider more disaggregated job categories. Reno is unique in that it experienced a very large decline in construction employment during the great recession. It is also unique because it is 9 miles from the Tahoe-Reno Industrial Center, home of Tesla's new Gigafactory and the world's largest industrial center. Thus, it may be important to separate out construction from other industrial sectors to analyze their impact. Another possible expansion of the current methodology would be to consider other demographic factors that could be related to different values for accessibility. Specifically, areas with higher concentration of employment of "working age" people, are likely to have higher demand from first-time homebuyers, while homes in areas with higher concentrations of younger adults may be more likely to be rentals. Another interesting approach would be to apply a Oaxaca Decomposition to the repeat sales in order to determine if there is evidence of changing preferences for job accessibility over time. For example, it is possible that during an economic downturn people may become more sensitive to transportation costs. Finally, it may be worthwhile to investigate the relationship of job access to foreclosure behavior. Intuitively, areas that lost more jobs during the recession are likely to have not only seen more volatile prices, but also to have been more susceptible to default.

References

William Alonso. *Location and land use: toward a general theory of land rent.* Harvard University Press, Cambridge, 1964.

Martin J Bailey, Richard F Muth, and Hugh O Nourse. A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304):933–942, 1963.

James D. Burnell. Industrial land use, externalities, and residential location. *Urban Studies*, 22(5):399–408, 1985.

Karl E Case and Robert J Shiller. The efficiency of the market for single-family homes. *The American Economic Review*, pages 125–137, 1989.

Assumpta Cerda and A El-Geneidy. Understanding the relationships between regional accessibility travel behaviour and home values. In *89th Annual Meeting of the Transportation Research Board, Washington, DC Transportation Research Board*, 2009.

John M Clapp, Mauricio Rodriguez, and R Kelley Pace. Residential land values and the decentralization of jobs. *The Journal of Real Estate Finance and Economics*, 22(1):43–61, 2001.

Eugene F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965. ISSN 00219398, 15375374. URL `http://www.jstor.org/stable/2350752`.

Joel P Franklin and Paul Waddell. A hedonic regression of home prices in king county, washington, using activity-specific accessibility measures. In *Proceedings of the Transportation Research Board 82nd Annual Meeting, Washington, DC*, 2003.

Dean H Gatzlaff and Donald R Haurin. Sample selection bias and repeat-sales index estimates. *The Journal of Real Estate Finance and Economics*, 14(1): 33–50, 1997.

Dean H Gatzlaff and Donald R Haurin. Sample selection and biases in local house value indices. *Journal of Urban Economics*, 43(2):199–222, 1998.

Kenneth Gibb, Liv Osland, and Gwilym Pryce. Describing inequalities in access to employment and the associated geography of wellbeing. *Urban Studies*, 51 (3):596–613, 2014.

Genevieve Giuliano and Kenneth A. Small. Subcenters in the los angeles region. *Regional Science and Urban Economics*, 21(2):163–182, 1991.

Genevieve Giuliano, Peter Gordon, Qisheng Pan, and JiYoung Park. Accessibility and residential land values: Some tests with new measures. *Urban Studies*, 47(14):3103–3130, 2010. doi: 10.1177/0042098009359949. URL `http://dx.doi.org/10.1177/0042098009359949`.

Edward L Glaeser, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer. Growth in cities. *Journal of political economy*, 100(6):1126–1152, 1992.

Walter G. Hansen. How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2):73–76, 1959. doi: 10.1080/01944365908978307. URL `http://dx.doi.org/10.1080/01944365908978307`.

James J Heckman. Sample selection bias as a specification error (with an application to the estimation of labor supply functions), 1977.

Sungsoon Hwang and Jean-Claude Thill. Influence of job accessibility on housing market processes: Study of spatial stationarity in the buffalo and seattle metropolitan areas. In *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, pages 373–391. Springer, 2010.

Michael Iacono and David Levinson. Location, regional accessibility, and price effects: Evidence from home sales in hennepin county, minnesota. *Transportation Research Record: Journal of the Transportation Research Board*, (2245): 87–94, 2011.

Michael Iacono and David Levinson. Accessibility dynamics and location premia: Do land values follow accessibility changes? *Urban Studies*, 54(2): 364–381, 2017. doi: 10.1177/0042098015595012. URL `http://dx.doi.org/10.1177/0042098015595012`.

Donald Jud and Terry Seaks. Sample selection bias in estimating housing sales prices. *Journal of Real Estate Research*, 9(3):289–298, 1994.

Paul Krugman. Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499, 1991. doi: 10.1086/261763. URL `https://doi.org/10.1086/261763`.

Bumsoo Lee. edge or edgeless cities? urban spatial structure in u.s. metropolitan areas, 1980 to 2000. *Journal of Regional Science*, 47(3):479–515, 2007.

Alfred Marshall. *Principles of Economics: An Introductory Volume*. 8th ed. London: MacMillan, 1920.

John William Matthews. The effect of proximity to commercial uses on residential prices. 2006.

Daniel P. McMillen. Identifying sub-centres using contiguity matrices. *Urban Studies*, 40(1):57–69, 2003a.

Daniel P McMillen. Neighborhood house price indexes in chicago: a fourier repeat sales approach. *Journal of Economic Geography*, 3(1):57–73, 2003b.

Daniel P McMillen. The return of centralization to chicago: using repeat sales to identify changes in house price distance gradients. *Regional Science and Urban Economics*, 33(3):287–304, 2003c.

Daniel P McMillen. Repeat sales as a matching estimator. *Real Estate Economics*, 40(4):745–773, 2012.

Daniel P. McMillen and John F. McDonald. Suburban subcenters and employment density in metropolitan chicago. *Journal of Urban Economics*, 43(2): 157–180, 1998.

Daniel P McMillen and Paul Thorsnes. Housing renovations and the quantile repeat-sales price index. *Real Estate Economics*, 34(4):567–584, 2006.

Jon P. Nelson. Accessibility and the value of time in commutingcommuting. *Southern Economic Journal*, 43(3):1321, 1977.

Liv Osland and Gwilym Pryce. Housing prices and multiple employment nodes: is the relationship nonmonotonic? *Housing Studies*, 27(8):1182–1208, 2012.

John R. Ottensmann, Seth Payton, and Joyce Man. Urban location and housing prices within a hedonic model. *Journal of Regional Analysis and Policy*, 38 (1):19–35, 2008.

A Small Kenneth, Song Shunfeng, et al. Population and employment densities: Structure and change. *Journal of Urban Economics*, 36(3):292–313, 1994.

Greg T Smersh and Marc T Smith. Accessibility changes and urban house price appreciation: a constrained optimization approach to determining distance effects. *Journal of Housing Economics*, 9(3):187–196, 2000.

Shunfeng Song. Modelling worker residence distribution in the los angeles region. *Urban Studies*, 31(9):1533–1544, 1994. doi: 10.1080/00420989420081411. URL http://dx.doi.org/10.1080/00420989420081411.

Shunfeng Song. Some tests of alternative accessibility measures: A population density approach. *Land Economics*, 72(4):474–482, 1996. ISSN 00237639. URL http://www.jstor.org/stable/3146910.

Johann Heinrich von Thunen. *The Isolated State*. Springer, 1826.

Nathan Wiseman. Selection bias in repeat sales and hedonic pricing models, 2017.

C. Y. Yiu and C. S. Tam. A review of recent empirical studies on property price gradients. *Journal of Real Estate Literature*, 12(3):307–322, 2004a.

CY Yiu and CS Tam. The estimation of housing price gradients: A comparison of different approaches applied in hong kong. In *International Conference on Adequate and Affordable Housing for All: Research, Policy, Practice, June*, pages 24–27. Citeseer, 2004b.